# Domain-Independent Detection of Emergency Situations Based on Social Activity Related to Geolocations

Hernan Sarmiento

Institute of Data Foundations
Department of Computer Science,University Of Chile
hsarmien@dcc.uchile.cl

**Abstract.** Most methods for detecting emergency situations using Twitter rely on identifying features within messages that contain domain-specific keywords. However, keyword-based methods require models to be trained on historical data of specific domains, in multiple languages, and for different types of events. In contrast, we propose to use self-organized geolocation related activity to identify emergency situations. To detect such events we track the frequencies, and probability distributions of the interarrival time of the messages related to specific locations. Using an off-the-shelf classifier that is independent of domain-specific features, we study and describe emergency situations based solely on location-based features in messages. Our findings indicate that anomalies in location-related social media user activity indeed provide information for automatically detecting emergency situations independent of their domain.

## 1 Introduction

During emergency situations, traditional media may suffer infrastructure issues and real-time communications could be disrupted. Instead, social network have played a critical role over the last fifteen years allowing users to share real-time information from people local to the incident, such as status updates, casualties, damages and alerts [9, 14].

Twitter is a microblogging platform that allows users to share short messages (called *tweets*) and is currently used worldwide by over 300 million people.[1] About 80% of Twitter users access from mobile devices, which contributes to the immediacy of diffusion of information, especially during crisis situations [5]. One of the most important tasks during these situations, is to timely detect incoming real-world events. In current works [4, 8, 10], these tasks are solved with methods that rely on keyword based filters on the *Twitter Public Streaming API*.[2] The problem with keyword-based methods is the need to train in specific domains for different type of events. For instance, Olteanu et al. [11] generate

[1] http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/
[2] https://developer.twitter.com/en/docs/tweets/filter-realtime/overview

a set of keywords based on different datasets. However, this is not sufficient for cases in which specific and previously unseen terms arise for a particular event (e.g., *#eqnz* for Earthquakes in New Zealand, or *#pabloph* for Typhoon Pablo in Philippines) [2, 13]. Furthermore, these sets of keywords are commonly obtained a particular language and will not work in others.

Here, we propose a method based on recurring references to country level locations in message metadata. For this task, we create a gazetteer tree based on the hierarchy of the specific country, divide the messages into fixed time-windows and compute the frequency and the probability distributions of the interarrival time of the messages for each geographical hierarchy. To detect an emergency situation, we train a SVM classifier by each hierarchy and apply a geographic spread to filter false positives detection considering that an emergency situation can be *Focalized* or *Diffused*.

Our main contribution is to create a methodology that detects instantaneous emergency situations just by using the location related information for a specific country. Furthermore, our classifiers do not depend on language to detect a new event since they do not use textual features as input. Moreover, we characterize crisis situations that affect small or large geographical areas.

The paper is organized as follows: we first introduce an overview of relevant literature related to emergency situation and event detection in social media. Next we present a complete description of our proposal divide in four parts. Thereafter, we provide a full description of our dataset and later we summarise our experimental validation over the ground truth and on-line evaluation. Finally, we deliver our discussions, conclusions and future work.

## 2 Related Work

Twitter has been used extensively during emergency situations to extract and identify relevant information. However, social media communications during emergency situations have become so abundant that it is necessary to sift through millions of data points to find information.

One of the main tasks related to emergency situations is to detect a new real-crisis event in social media. Most existing event-detection methods described in the literature are based on keywords. For instance, *EMERSE* [4] used a set of keywords related to the Haiti earthquake and applies a random forests algorithm for classifying messages. Likewise, the *Twicalli* system [10] introduced an unsupervised approach to detect earthquakes that only requires a general list of keywords.

Researchers at CSIRO Australia proposed *ESA* [3], a system to detect disasters in Australia and New Zealand. This system is based on a probabilistic method to identify bursty keywords and historical data to build a language model of word occurrences. Alerts are identified if a term has a probability distribution that significantly deviates from the language model. Similarly, *Twitcident* [1] system was developed for detecting incidents which relies on emergency broadcasting services, such as the police, the fire department and other public emergency services. The *Twitcident* framework translates the broadcasted message into an initial incident profile that is applied as query to collect messages from Twitter, where an incident profile is a set of weighted attribute-value pairs that describe the characteristics related to the incident.

## 3 Proposed Approach

In order to provide a complete coverage of location-based detection of emergency situations, we propose an approach which has four stages. Next, we describe each stage:

### 3.1 Data Pre-Processing

Since we consider users as citizen sensors, we filter messages depending of native language of each country using the attribute *lang* retrieved from the metadata in tweets[3]. We remove user mentions, URLs, special characters and apply text tokenisation. Also we do not remove hashtags or stopwords because some locations can be included in hashtags terms and some places can contain stopwords that can differentiate them from other similar non-related to locations terms.

### 3.2 Signal Creation

The problems that we address in this stage are how to infer locations from microblog messages to track signals over time in the data stream and what is the lowest hierarchy level to find and assign to these signals.

**Geographical Hierarchy** Using the idea of *Gazetteer as a Tree* presented in [16] in which each place is associated with a canonical taxonomy node, we construct our gazetteer tree based on Geonames[4] and Wikipedia[5] for each country to analyse. In our approach, we use a subset of the gazetteer hierarchy with three levels, in which the lowest level is represented by a city since a large amount of users specify their location down to this level [7]. Furthermore, the name of locations are considered just in the native language of the country.

**Finding Locations in Microblog Metadata** Considering the geographical hierarchy explained in the previous section, we search these locations on different levels of metadata and create one signal for each as following: *tweet text*, where the location is mentioned on the body of the message; *user location*, where the location is mentioned the location settled by the user in their profile; and *tweet text - user location*, where the location is mentioned in the body of the message and the user who shares message has the same location in his profile.

---

[3] https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object
[4] http://download.geonames.org/export/dump/
[5] http://www.wikipedia.org/

### 3.3 Time-Windows

In this stage we address the problems of how to divide and determine the time-window size to detect a new emergency situation and what features by the time-window allow it.

**Determining Optimal Window Size** According to the works of Guzman and Poblete, and Maldonado et al. [6, 10] we divide our signals into windows of six minutes because it divides a 24-hour day exactly and decrease the occurrence of empty windows for a term, making the analysis easier to understand and to compare from different days.

**Normalized Frequency** We compute the number of the messages of each time-window by signal. To normalize frequency, we compute *z-score* as following:

$$zscore = \frac{x_i - \mu_k}{\sigma_k} \tag{1}$$

where $x_i$ is the frequency of the current $i$ time-window, $\mu_k$ and $\sigma_k$ are mean and standard deviation of the previous $k$ time-windows respectively.

**Interarrival Time** To characterize the urgency of the messages during a time-window, we compute the *interarrival time* which is defined as $d_i = t_{i+1} - t_i$, where $d_i$ denotes the difference between two consecutive social media messages $i$ and $i+1$ that arrived in moments $t_i$ and $t_{i+1}$ respectively.

To quantify a high-frequency in very small values of $d_i$, we compute the measures *skewness* and *kurtosis*, which represent the asymmetry and the tailedness of the shape of probability distribution respectively. Finally, we apply the equation 1 over *skewness* and *kurtosis* to calculate variation based on previous values.

### 3.4 Geographic Spread

An emergency situation that affects and mobilizes response in a small area is defined as *focalized*, while a disaster with a large geographic impact is defined as *diffused* [12]. Using this definition, we extend this concept to represent neighbourhoods between locations obtained from section 3.2. For that purpose, we create an *adjancency matrix* $M$, where $M_{i,j} = 1$ represents if two locations are geographically connected and $M_{i,j} = 0$ if they are not connected. For instance, if an event is diffused (e.g., earthquake), the detection should be in adjacent-locations independently of metadata-level. On the other hand, if an event is focalized (e.g., terrorist attack), just one location should be detected but in different metadata-levels simultaneously.

## 4 Dataset

Data is collected from Twitter Public Streaming API, which allows access to subsets equal to 1% of public random status descriptions in real-time. In our approach, we get entire subsets of messages without use keywords or specific locations.

In this work, our ground truth were five earthquakes with magnitudes between $5.5Mw$ and $7.6Mw$, occurring in Italian-speaking and Spanish-speaking countries between October 2016 and April 2017. For that purpose, we collected 20 million of messages 12-hours before and after the emergency situation events. According our proposal we create both the gazetteer hierarchies[6] for each country and construct the signals based on each hierarchy and metadata-level (Table 1).

**Table 1.** Number of messages by signal.

| Hierarchy | Metadata-level | Messages |
|---|---|---|
| All | All | 87,291 |
| Country | Tweet Text | 11,584 |
| | User Location | 25,313 |
| | Tweet Text - User Location | 1,417 |
| State | Tweet Text | 4,110 |
| | User Location | 13,352 |
| | Tweet Text - User Location | 86 |
| City | Tweet Text | 1,415 |
| | User Location | 8,971 |
| | Tweet Text - User Location | 20 |

As a result of the number of messages of each signal , we discard all signals related to city hierarchy since a great amount of small cities have zero frequency in a normal situation unlike to capital or metropolitan cities.

The exactly datetime event is obtained from National Seismology Agency in Chile[7] and National Institute of Geophysics and Volcanology in Italy[8]. With the purpose of labeling a time-window as positive class (detection), we set as detection those time-windows with positive variation in frequency, skewness and kurtosis with respect to the normalization of the previous values. Moreover and according to (Figure 1), we include the three next time-windows after the event to compensate the imbalance between classes given that after these number of time-windows, the variation in the features decrease.

[6] http://users.dcc.uchile.cl/∼hsarmien/gazetteer.html
[7] http://www.sismologia.cl/
[8] http://www.ingv.it/it/

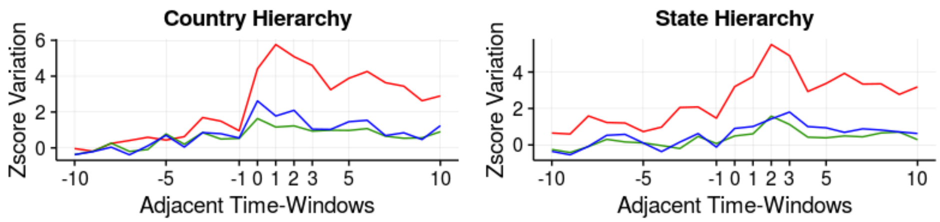


**Fig. 1.** Average variation in emergency situations between time-windows. Red line represents normalized frequency, green line normalized skewness and blue line normalized kurtosis.

## 5 Experiments

Our filtering task can be seen as binary classification task. For this reason, we employed traditional binary classifier Support Vector Machine *(SVM)* and we separated country and state in different datasets and set both kernels and classification parameters independently.

Given that an emergency situation is not an usual event, we have an highly unbalanced data respect to the classes after labeled (1 ≈ 2% of positive class corresponding to *detection*). Therefore, we used *under-sampling* over country and state datasets increasing our possitive class to 15 ≈ 18%. Additionally, to validate our model, we used *5-fold cross-validation* where one earthquake dataset is used as testing and the remaining earthquakes dataset as training.

**Independent Analysis of Hierarchies** Our first analysis is just considering the hierarchies as isolated detections. The top of the Table 2 shows the results considers only the prediction over each label in our datasets. Low values in the Precision and FPR can be explained by multiple *bursts* in our signals which do not correspond to emergency events.

In addition to the analysis of number of detections by labels, we also study the number of detections by time-windows. For this analysis, we search the time-windows for each hierarchy where the all metadata-levels are well classified with correct class. According to the results shown on the middle of the Table 2, when we analyse country and state independently the values of Precision, F1 and FPR have worst values than the analysis by label .

**Dependent Analysis of Hierarchies** Our second analysis considered the hierarchies as non-isolated detections. We inspected the time-windows where all metadata-level for country and state hierarchy have a correct detection simultaneously. The results are shown in the row *Country-State* in the Table 2. In contrast to the independent analysis of country and state, we improved the Precision, F1 and FPR values as a consequence of a smaller amount of the time-windows related to non emergency situations are assigned as detection. However,

**Table 2.** Average performance of 5-fold cross-validation by hierarchy and geographic spread (G.S.)

| | Hierarchy | P | R | F1 | FPR |
|---|---|---|---|---|---|
| label | Country | 0.3 | 0.83 | 0.45 | 0.14 |
| | State | 0.35 | 0.83 | 0.5 | 0.08 |
| time-window | Country | 0.15 | 0.77 | 0.25 | 0.15 |
| | State | 0.17 | 0.88 | 0.29 | 0.12 |
| | Country-State | 0.35 | 0.7 | 0.47 | 0.03 |
| | Country(2)-State with G.S. | 1 | 0.64 | 0.78 | 0 |
| | Country(3)-State with G.S. | 1 | 0.47 | 0.64 | 0 |

when we see the value obtained for FPR ($FPR = 0.03$), this rate represents an incorrect number of time-windows assigned as detection equal to 23. This means that we have 23 new emergency situations detected by our classifier.

**Geopraphic Spread Analysis** Our third analysis considered the hierarchies as non-isolated detections and applies the Geographic Spread (G.S.). We considered as a correct detection those time-windows where the state/s classified as detection are defined as *Focalized* or *Diffused* and exist dependency between hierarchies.

In addition to the results of the dependency analysis explained above, we see that a large amount of time-windows for country hierarchy ($\approx 82\%$) have more than one metadata-level when exist a correct detection.

Considering the geographic spread by states and the number of metadata-levels by country hierarchy, we analysed the results shown on the bottom of the Table 2. On the one hand, the row with value equal to *Country(2)-State with G.S.* represents the detection when we considered at least two metadata-levels for the country hierarchy and the geographic spread for states. In contrast to the the previous analyses, we improved the values of the Precision, F1 and FPR. The last metric is very important because there are no time-windows incorrectly assigned as emergency situations. Consequently, the Recall values decrease which means that our method remove some time-windows classified as detection. Beside the percent of emergency situations detected is equal to 100% with a average delay equal to 10.4 minutes from the impact of the event to the first detection.

On the other hand, *Country(3)-State with G.S.* represent the detection when we considered three metadata-levels for country hierarchy and the geographic spread for states. Similar to *Country(2)-State with G.S.*, we improved the values of Precision, F1 and FPR but our recall decrease from $R = 0.58$ to $R = 0.47$, detecting 80% of the emergency situations with a average delay equal to 11.5 minutes from the impact of the event to the first detection.

### 5.1 Online Evaluation

For our evaluation in the Twitter Public Stream, we trained classifier with five earthquakes identified in our ground truth. Furthermore, our evaluation dataset is formed by seven different events that occurred in England between December 2016 and October 2017. For each event we considered the full-day in which they occurred. Geographic spread analysis is used to evaluate our method because decrease the number of false positives detection (Table 3 and Table 4).

**Table 3.** Online evaluation by time-windows using Country(2)-State with G.S. method

| Event | Detected | Before Event | After Event | Delay (min) | Top 3 Bigrams |
|---|---|---|---|---|---|
| Premier League Soccer Matches | 2 | - | - | - | (man, utd), (new, year), (happy, new) |
| Westminster Terrorist Attack | 13 | 0 | 13 | 32 | (stay, safe), (terror, attack), (safe, everyone) |
| Manchester Terrorist Attack | 12 | 1 | 11 | 23 | (ariana, grande), (incident, arena), (grande, concert) |
| London Terrorist Attack | 14 | 7 | 7 | 36 | (stay, safe), (incident, bridge), (borough, market) |
| U.K. Elections | 5 | - | - | - | (theresa, may), (vote, labour), (van, dijk) |
| England vs Slovenia Soccer Match | 4 | 4 | 0 | - | (simon, brodkin), (join, us), (theresa, may) |
| Metallica Live in London | 4 | 4 | 0 | - | (always, said), (chance, win), (carabao, cup) |

The first evaluation *Country(2)-State with G.S.* has full detection of the terrorist attacks with average delay time equal to 30.3 minutes. These detections are related to the event given that the bigrams represent terms associated with crisis situations. However, the London Terrorist Attack has 50% of the detected time-windows after the event, which means that there are seven time-windows non-related to emergency situations. Besides the crisis situations analysis, we also study the number of detected time-windows in non-related to emergency situation events. In the same way, we have a large amount of misclassified time-windows that do not represent crisis situations.

The second evaluation *Country(3)-State with G.S.* decreases the number non-related to emergency situations events detected as crisis situations. We can see three time-windows in two events detected as emergency situations. Furthermore, when we analysed the number of the detected emergency situations, two-thirds (66%) of the events are detected correctly with average delay time equal to 30.3 minutes. In the case of London Terrorist Attack, our method detects one time-

**Table 4.** Online evaluation by time-windows using Country(3)-State with G.S. method

| Event | Detected | Before Event | After Event | Delay (min) | Top 3 Bigrams |
|---|---|---|---|---|---|
| Premier League Soccer Matches | 0 | - | - | - | |
| Westminster Terrorist Attack | 4 | 0 | 4 | 32 | (terror, attack), (stay, safe), (terrorist, attack) |
| Manchester Terrorist Attack | 2 | 0 | 2 | 23 | (ariana, grande), (praying, everyone), (everyone, affected) |
| London Terrorist Attack | 1 | 1 | 0 | - | (ariana, grande), (around, world), (lady, gaga) |
| U.K. Elections | 0 | - | - | - | |
| England vs Slovenia Soccer Match | 1 | 1 | 0 | - | (per, day), (menswear, sample), (closed, roads) |
| Metallica Live in London | 2 | 2 | 0 | - | (happy, birthday), (chance, win), (always, said) |

window before the event but the bigrams describe that the detections do not correspond to crisis situations.

## 6 Discussion

Our findings suggest that there is evidence to detect an emergency situation based on anomaly frequency of messages that contain locations for a specific country. Indeed, our method based on the number of metadata-levels by country hierarchy and geographic spread by state, detects a 80% of the events related to emergency situations as we could demonstrate in our ground truth. Also, our method is independent of the textual features because we apply the model over different languages as Spanish, Italian and English. Furthermore, we test our model in different types of events such as earthquakes and terrorist attacks.

However, when we apply our method in the online evaluation, we detect 66% of the emergency situations that affected England. This explains that the signals, and for various reasons: the number of active users in United Kingdom[9] which can affect the anomaly frequency of the messages since there exists a high daily average activity of the messages; similar locations in other countries (York $\approx$ New York); and soccer teams with names of cities (Manchester United, Liverpool). These issues also can affect the number of false positive detection in which in the case of England was 30% of the non-related to emergency events.

[9] http://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ visited on January 2018

Regarding the geographic spread where we define an emergency situation as diffused or focalized, we find some evidence that differentiates them. In the case of diffused events, the delay time of the our first detection was less than 12 minutes and in focalized events was greater than 30 minutes (Figure 2). This explains that, in diffused events such as earthquakes, a high number of people are affected at the same time by an event which generates a collective reaction in social media in the locations where the event impacted. In Figure 2, we can see that earthquakes have at least two detected locations in the first detection (except Italy EQ2). In contrast, focalized events have less amount of eyewitness, then when the users share messages in social media, the frequency does not affect the average daily message of the country in the first minutes. This can be explained in Figure 2 where the terrorist attacks have just 1 detected location in the first detection.

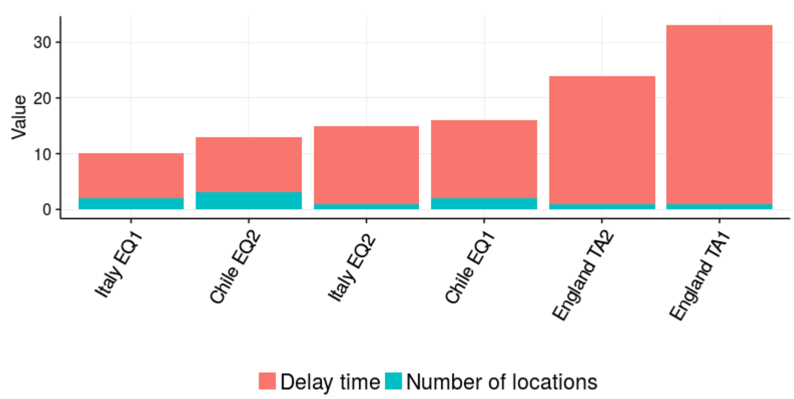


**Fig. 2.** Delay time and number of locations in the first detection for diffused and focalized emergency situations.

## 7 Conclusion

In this paper we have presented a methodology for detecting an emergency situation based on location for a specific country. This approach is independent of the textual features and can be used in different types of events and languages. We furthermore have presented an analysis of geographic spread for different types of events that can be categorized. However, our experiment considers just a small portion of emergency situations, which is not representative for all types of crisis situations.

In our future work, we will add Point of Interest to our gazetteer tree to increase the frequency by time-windows in each hierarchy. We also plan to study the relevance of the different metadata-levels and assign weights for each. Finally, we will create a web application to visualize events in real-time.

# Bibliography


[1] Fabian Abel, Claudia Hauff, Geert-Jan Houben, Richard Stronkman, and Ke Tao. 2012. Twitcident: fighting fire with information from social web streams. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, 305–308.

[2] Axel Bruns and Jean E Burgess. 2012. Local and global responses to disaster:# eqnz and the Christchurch earthquake. In *Disaster and emergency management conference, conference proceedings*, Vol. 2012. AST Management Pty Ltd, 86–103.

[3] Mark A Cameron, Robert Power, Bella Robinson, and Jie Yin. 2012. Emergency situation awareness from twitter for crisis management. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, 695–698.

[4] Cornelia Caragea, Nathan McNeese, Anuj Jaiswal, Greg Traylor, Hyun-Woo Kim, Prasenjit Mitra, Dinghao Wu, Andrea H Tapia, Lee Giles, Bernard J Jansen, et al. 2011. Classifying text messages for the Haiti earthquake. In *Proceedings of the 8th international conference on information systems for crisis response and management (ISCRAM2011)*. Citeseer.

[5] Carlos Castillo. 2016. *Big crisis data: social media in disasters and time-critical situations*. Cambridge University Press.

[6] Jheser Guzman and Barbara Poblete. 2013. On-line relevant anomaly detection in the Twitter stream: an efficient bursty keyword detection model. In *Proceedings of the acm sigkdd workshop on outlier detection and description*. ACM, 31–39.

[7] Brent Hecht, Lichan Hong, Bongwon Suh, and Ed H Chi. 2011. Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 237–246.

[8] Muhammad Imran, Carlos Castillo, Ji Lucas, Patrick Meier, and Sarah Vieweg. 2014. Aidr: Artificial intelligence for disaster response. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*. International World Wide Web Conferences Steering Committee, 159–162.

[9] Muhammad Imran, Shady Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. Extracting information nuggets from disaster-Related messages in social media.. In *ISCRAM*.

[10] Jazmine Maldonado, Jheser Guzman, and Barbara Poblete. 2017. A Lightweight and Real-Time Worldwide Earthquake Detection and Monitoring System Based on Citizen Sensors. In *Proceedings of the Fifth Conference of Human Computation and Crowdsourcing*. AAAI, 137–146.

[11] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2014. CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises.. In *ICWSM*.

[12] Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. 2015. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 994–1009.

[13] Liza Potts, Joyce Seitzinger, Dave Jones, and Angela Harrison. 2011. Tweeting disaster: hashtag constructions and collisions. In *Proceedings of the 29th ACM international conference on Design of communication*. ACM, 235–240.

[14] Kevin Stowe, Michael J Paul, Martha Palmer, Leysia Palen, and Kenneth Anderson. 2016. Identifying and categorizing disaster-related tweets. In *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*. 1–6.

[15] Maximilian Walther and Michael Kaisser. 2013. Geo-spatial Event Detection in the Twitter Stream.. In *ECIR*. Springer, 356–367.

[16] Jie Yin, Sarvnaz Karimi, and John Lingad. 2014. Pinpointing locational focus in microblogs. In *Proceedings of the 2014 Australasian Document Computing Symposium*. ACM, 66.